Int. J. Human-Computer Studies 82 (2015) 57–68

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Utilizing linguistically enhanced keystroke dynamics to predict typist cognition and demographics ☆

David Guy Brizan [a], Adam Goodkind [a], Patrick Koch [b], Kiran Balagani [b], Vir V. Phoha [c], Andrew Rosenberg [a,d,*]

[a] CUNY Graduate Center, New York, NY 10016, USA
[b] New York Institute of Technology, Old Westbury, NY 11568, USA
[c] Louisiana Tech University, Ruston, LA 71272, USA
[d] Computer Science Department, CUNY Queens College, Queens, NY 11367, USA

ARTICLE INFO

Article history:
Received 11 August 2014
Received in revised form
23 April 2015
Accepted 27 April 2015
Communicated by Keith Vertanen
Available online 21 May 2015

Keywords:
Keystroke dynamics
Stylometry
Cognitive load recognition
Demography recognition
Typing production

ABSTRACT

Entering information on a computer keyboard is a ubiquitous mode of expression and communication. We investigate whether typing behavior is connected to two factors: the cognitive demands of a given task and the demographic features of the typist. We utilize features based on keystroke dynamics, stylometry, and "language production", which are novel hybrid features that capture the dynamics of a typists linguistic choices. Our study takes advantage of a large data set (∼350 subjects) made up of relatively short samples (∼450 characters) of free text. Experiments show that these features can recognize the cognitive demands of task that an unseen typist is engaged in, and can classify his or her demographics with better than chance accuracy. We correctly distinguish High vs. Low cognitively demanding tasks with accuracy up to 72.39%. Detection of non-native speakers of English is achieved with $F_1 = 0.462$ over a baseline of 0.166, while detection of female typists reaches $F_1 = 0.524$ over a baseline of 0.442. Recognition of left-handed typists achieves $F_1 = 0.223$ over a baseline of 0.100. Further analyses reveal that novel relationships exist between language production as manifested through typing behavior, and both cognitive and demographic factors.

1. Introduction

As early as World War II, U.S. military intelligence began to identify individuals by their rhythms in tapping out Morse code messages. These rhythms, called "the Fist of the Sender," supported the tracking of Morse code operators and therefore the troops and vehicles moving with the operators. Since then, keystroke dynamics or typing dynamics have had a number of practical applications, including the determination of the cognitive demands of individuals (Vizer et al., 2009). In much the same way individual operators may be identified by their tapping rhythms and individual speakers may be identified through spectral and prosodic features of their speech (Reynolds et al., 2000; Shriberg et al., 2004), individual typists exhibit unique but self-consistent typing patterns (Gunetti, 2005; Sheng et al., 2005).

A related field, stylometry, is also concerned with author identification (Canales et al., 2011). Stylometry describes the measurement of linguistic "style" and has been effectively used in authorship attribution (Juola, 2006; Stamatatos, 2009), in dating a single piece of writing (Can and Patton, 2004) and in establishing genre shifts within the work of a single author (Kessler, 1997). However, whereas keystroke dynamics has been used to verify the identity of one of hundreds of typists, stylometric applications typically distinguish between many fewer individuals.The metrics developed within this field typically rely on the user's spelling of specific words, choice of words in a sentence and choices with respect to grammar. Stylometric analyses are applied to prepared, static text.

In this paper, we describe two applications of combining keystroke dynamics, stylometry and a new set of language production features: to identify the type of cognitive task a typist is performing and to identify three demographic cohort identifications. We collect data from subjects' typed responses to prompts which require them to engage in one type of task or another. This analysis is non-interruptive and non-intrusive. While our experiments operate on data collected from local computer users, our work could easily be extended to remote users via keystroke logging software.

In the case of predicting the type of cognitive task, we aim to determine whether the user is performing a cognitively simple task,

☆This paper has been recommended for acceptance by Keith Vertanen.
* Correspondence to: Computer Science Department, Queens College, 65-30 Kissena Blvd., Queens, NY 11367, United States. Tel.: +1 718 997 3562.
E-mail addresses: dbrizan@gradcenter.cuny.edu (D.G. Brizan), agoodkind@gradcenter.cuny.com (A. Goodkind), pkoch03@nyit.edu (P. Koch), kbalagan@nyit.edu (K. Balagani), phoha@latech.edu (V.V. Phoha), andrew@cs.qc.cuny.edu (A. Rosenberg).
URL: http://eniac.cs.qc.cuny.edu/andrew (A. Rosenberg).

http://dx.doi.org/10.1016/j.ijhcs.2015.04.005
1071-5819/

such as recalling known information, or performing a more cognitively taxing task such as analyzing an argument or creating a new idea. The research presented in this paper rests on two assumptions. First, we assume that performing differing tasks will have differing associated cognitive demands. This assumption is based on the Bloom Taxonomy (Anderson et al., 2001), widely used in education to categorize the cognitive demands of instructional activities. Second, we assume that the cognitive activity of a typist, particularly when performing a language production task, is reflected in his or her typing behavior. This assumption is supported by the findings of Vizer et al. (2009), which observed that a user's typing patterns vary based on the cognitive demands of a task. Specifically, we hypothesize that the cognitive demands of performing a task will have an observable impact on a typist's behavior that can be measured through features related to keystroke dynamics, stylometry, and language production. In the first set of experiments presented in this paper, subjects respond to prompts which are drawn from different types of cognitive tasks with varying complexity. We then predict the type of task a subject is performing given the typing patterns and final response.

For demographic prediction, we divide our subjects along three broad demographic dimensions: gender, dominant hand and primary language (native vs. non-native speakers of English). Each of these demographic divisions may be viewed as a cohort with a different set of keystroke dynamics when compared to its counterpart. We aim to be able to place a user in a cohort based on the user's typing patterns and language use, such as "left-handed, female, native English-speaker". In the context of user identification and verification, this can be used as a filter to eliminate some candidates from further consideration enabling more focused downstream analysis.

This work employs a number of novel features for keystroke dynamics and stylometry. In addition to measuring hold and interval times of each key individually, we explore aggregations of keys based on their keyboard position, which distinguishes, for example, keys typed by the left and right hand. By performing stylometric analysis on streams of typed data, we are able to develop features measuring revision behavior in addition to the final, static text. Moreover we develop a number of language production features which extend traditional stylometric measures with information about their timing.

The most important contributions of our study are

- We demonstrate how the type of task a typist is performing—based on the expected cognitive demand—affects typing output. Previous studies have centered around a homogeneous task type, whereas we can show the effects of varying cognitive demands.
- We propose and implement a new class of features, keystroke language production. These features take advantage of both keystroke dynamics and stylometry, to capture the dynamics, or prosody, of a typist's language production.
- The text being analyzed in this work is entered freely, with minimal constraints as to length or content. Moreover, predictions are made using much less data per answer than comparable studies, and with significantly more subjects.
- Typical studies of this kind (cf. Section 2) attempt to model the behavior of a typist and compare subsequent samples of the same person's typing to this model. In this work, we demonstrate the value of typing behavior to generalize to unknown typists, i.e. those not seen during training.

Our paper is structured as follows: Section 2 discusses related work and applications. Section 3 describes the methods that are in common between the two sets of experiments including details of the data collection (Section 3.1) and a description of the features we analyze (Section 3.2). Sections 4.1 and 4.2 describe experiments in predicting cognitive task and demography from an unknown typist, respectively. We conclude and discuss future work in Section 5.

2. Related work

Many researchers have employed keystroke dynamics for a variety of purposes including individual and cohort identification, identification of the typist's stress level or emotional state and a measurement on cognitive performance (Epp et al., 2011; Monrose and Rubin, 2000; Bergadano et al., 2003). Because keystroke dynamics uses available hardware common to most computer systems, it is especially effective as a "soft biometric" in authentication systems (Bartlow and Cukic, 2006; Villani et al., 2006; Joyce and Gupta, 1990). A "soft biometric"-predicts demographic classification but not specific identity.

Specifically with respect to the cognitive complexity of a task, Thomas et al. (2005) found a negative correlation between programming performance and typing speed. While programming is a cognitively demanding task, this experiment was designed for a specific population group (student programmers) performing a specific task and the specific task is essentially making the experiment a speed-accuracy tradeoff in a limited domain.

In work that serves as support for our assumption that cognitive demands have an impact on typing behavior, Vizer et al. (2009) used features drawn from keystroke dynamics (as well as stylometry) to infer levels of cognitive stress with 75% accuracy. This work discusses unobtrusive monitoring of physical and mental health, for example in the aging population. The experiments in Vizer et al. (2009) require each subject register with the system and establish performance means and ranges under different stress conditions. These distributions are later used to determine the stress of the subjects by comparing to a recorded set of distributions.

For gender prediction, both Fairhurst and Costa-Abreu (2011) and Giot and Rosenberger (2012) found traction in predicting gender with the use of features from keystroke dynamics; in fact, they further employed this prediction as a soft biometric in user identification on the GREYC data set (Giot et al., 2009). Fairhurst and Costa-Abreu (2011) reports approximately 80% accuracy in determining gender with a single classifier and (Giot and Rosenberger, 2012) reports 91.63% accuracy, exceeding the 73% baseline. However, their experiments require the user to type pre-defined text (specifically a password string) on which the determination of gender is performed.

Detecting a typist's dominant hand (handedness) is an intuitive application of keystroke dynamics. Some of the earliest mentions of using keystroke dynamics for authentication (Monrose and Rubin, 1997) note this. Experiments in the prediction of handedness were undertaken in Idrus et al. (2014), though they were tested on a fixed set of passwords, with the same users employed for training and testing.

Stylometric features have been used in predicting gender differences in writing styles with success (Goswami, 2009; de Vel et al., 2002; Koppel et al., 2002). Koppel et al. (2002) report 77.39% accuracy (on a 49.4% baseline with a minimum of 500 words in the training set), 70.2% F_1 (on a minimum of 150 training words) and an accuracy range of "approximately 80%". Similar to work presented here, de Vel et al. (2002) and Koppel et al. (2002) extract an extensive feature set without assumptions about what stylometric features would be most useful.

Our experimental design is closest to Sarawgi et al. (2011), in which the topics and genres have been carefully balanced to avoid bias. Similar to our results, they find character-level features are best at differentiating between males and females, yielding 71.3% accuracy in doing so with a training set composed of 430–470 words per user. Bergsma et al. (2012) also use stylometric features to predict gender (at 48.2% F_1 score) as well as native language